

Rishu Kumar

✉ kumar@ufal.mff.cuni.cz | 🏠 rishu.uk | 🌐 pyRis

Education

Charles University + Saarland University

Prague, CR | Saarland, DE

EUROPEAN MASTER'S PROGRAMME IN LANGUAGE & COMMUNICATION TECHNOLOGIES

(2 years) 2021/22 - 2022/2023 (exp.)

• Projects:

- Training interactive Dialogue Systems from Scratch with DailyDialog and MultiWOZ dataset,
- Pre-training BERT on synthetic text for Machine Translation for $EN \leftrightarrow \{DE, CZ\}$,
- Summarization System for context retention in generating theatre script with customized GPT-2 beyond 1024 tokens, for THEaiTRE project (in progress)

Bhagalpur College of Engineering

Bhagalpur, India

B.TECH IN COMPUTER SCIENCE ENGINEERING

(4 years) 2014/15 - 2017/18

- **Major Project:** *e-Dustbin: A fully automated dustbin using Arduino to optimize route for waste pickup vehicles*
- **Summer School:** Indian Institute of Science, Bangalore

Publications

Team ÚFAL @ CreativeSumm: BART and SamSum based few-shot approach for creative Summarization

COLING'22

Rishu Kumar, Rudolf Rosa

Team ÚFAL at CMCL 2022 Shared Task: Figuring out the correct recipe for predicting Eye-Tracking features using Pretrained Language Models

CMCL (ACL 2022)

Sunit Bhattacharya, Rishu Kumar, Ondřej Bojar

ELITR multilingual live subtitling: Demo and strategy

EACL 2021

Ondřej Bojar, Dominik Macháček, Sangeet Sagar, Otakar Smrž, Jonáš Kratochvíl, Peter Polák, Ebrahim Ansari, Mohammad Mahmoudi, Rishu Kumar, Dario Franceschini, Chiara Canton, Ivan Simonini, Thai-Son Nguyen, Felix Schneider, Sebastian Stüker, Alex Waibel, Barry Haddow, Rico Sennrich, Philip Williams

Operating a Complex SLT System with Speakers and Human Interpreters

MtSummit 2021

Ondřej Bojar, Vojtěch Srdečný, Rishu Kumar, Otakar Smrž, Felix Schneider, Barry Haddow, Phil Williams, Chiara Canton

SEntFiN 1.0: Entity-Aware Sentiment Analysis for Financial News

JASIST

Ankur Sinha, Satishwar Kedas, Rishu Kumar, Pekka Malo

Short-Term Word-Learning in a Dynamically Changing Environment

in review(EMNLP)

Christian Huber, Rishu Kumar, Ondřej Bojar, Alexander Waibel

ELITR Deliverables

ELITR

Paper on Benchmarks for Minuting

in review

Main Author

Paper on using Minuting for coherent play script generation

in writing

Main Author

Work Experience

Saarland University

Spoken Language Systems

TUTOR

Oct 2022 - Current

- Tutoring Neural Networks Theory and Implementations

LEXITAS

United States

RESEARCH INTERN

Jun 2022 - Current

- Implemented clustering and semantics similarity algorithms to find similar questions
- Contributed in topical and abstractive summarization of legal transcripts
- Incorporated concepts from dialogue systems in discourse analysis

European Live Translator (Charles University)

Institute of Formal and Applied
Linguistics

Aug 2020 - Mar 2022

RESEARCH INTERN

- Preparing pipeline for continuous evaluation of ELITR-testset
- Improving SLTev a tool to evaluate ASR, MT and SLT performance
- Domain adaptation / retraining of ASR models (English and Czech) to improve quality and subsequently produce better translation
- Creating and maintaining bash pipelines to connect ASR, MT and publishing workers
- Stress testing of systems to ensure reliability before using in research demo at multiple venues
- Contributed in converting offline models to online models and creating Docker images for the same
- Created pipelines for automation of real-time meeting summarization of meeting in collaboration with alfaview and PerVoice

Indian Institute of Management Ahmedabad

Production & Quantitative Methods

Mar. 2018 - Jul. 2020

RESEARCH ASSISTANT

- Developed lexicon-based language models capable of extracting entity-based sentiments; Average news headline length ~ 8 words
- Released a annotated dataset of 12,000+ financial news headlines
- Compared GBMs, SVMs, and MLPs for classification; Contrasted feature-engineered input vectors vs GloVe and Word2Vec Achieved ~ 91% accuracy and ~ 93% F1-score using RoBERTa
- Validated the effect of sentiment information contained in news-flow on over-night aggregate market (NSE 500 index) returns
- Audited a PhD level course titled Non-Linear Optimization

Vidhi Centre for Legal Policy

FREELANCE DATA SCIENTIST

Sep. 2019 - Oct. 2019

- A first-of-its-kind study on the evaluation of judicial performance using a novel methodology: Case study of the Delhi High Court
- Developed web scrapers to extract judicial information and judgment data using Python libraries
- Implemented data-processing pipelines using regular expressions to extract information from text and store it in a tabular format

Academic Misc

Reviewing Work and Organization

- Co-organizing Low-Res Machine Translation shared task @ IWSLT 2023
- Reviewer for Workshop on Information Extraction from Scientific Publications @ ACL-IJCNLP 2022
- Reviewer for Generation, Evaluation & Metrics (GEM) Workshop @ EMNLP 2022

Others

- Elected as Student Representative for European Master's Programme in Language Science and Technologies for 2021-22
- Secured a rank of 3,978 in Graduate Aptitude Test in Engineering'19 among 99,932 candidates
- Started the first coding club in my college and organized competitions to promote it
- One of 72 selected applicants for an Undergraduate Summer School at Computer Science and Automation Dept., Indian Institute of Science, Bangalore; Only selected candidate from Bihar
- Worked as a core team member for annual Techno-Management festival in college; Worked as a facilitator between tech-team and sponsorship team; Designed the logo for college festival
- All India Rank 419 in IndiaHacks'16, an algorithm challenge on HackerEarth; College Rank 1

Technical Knowledge

Programming Python, C/C++, bash, Matlab, HTML5

Frameworks PyTorch, TensorFlow, Keras, Scikit-learn, Kaldi, MySQL, Marian

Miscellaneous Latex, Selenium, MS Excel, SGE, Unix

Languages

Hindi Native

English C2 (iBT TOEFL 112/120)

Czech A2 (in development)

German A1 (in development)